López-Labraca, Javier, Fernández-Torres, Miguel ángel, González-Díaz, Iván, Díaz-de-María, Fernando, Pizarro, Ángel. (2018) Enriched dermoscopic-structure-based cad system for melanoma diagnosis. *Multimedia Tools and Applications*, 77(10), pp.: 12171-12202.

DOI: https://doi.org/10.1007/s11042-017-4879-3

Noname manuscript No.
(will be inserted by the editor)

Enriched Dermoscopic-Structure-Based CAD System for Melanoma Diagnosis

Javier López-Labraca · Miguel Ángel Fernández-Torres · Iván González-Díaz · Fernando Díaz-de-María · Ángel Pizarro.

Received: date / Accepted: date

Abstract Computer-Aided Diagnosis (CAD) systems for melanoma detection have received a lot of attention during the last decades because of the utmost importance of detecting this type of skin cancer in its early stages. However, despite of the many research efforts devoted to this matter, these systems are not used yet in everyday clinical practice. Very likely, this is due to two main reasons: 1) the accuracy of the systems is not high enough; and 2) they simply provide a parallel diagnosis that actually does not help to the doctors (as long as there is no way to interpret it).

In this paper, we propose a novel approach that aims to provide the doctor with an enriched diagnosis. Specifically, we rely on a dermoscopic-structure-based soft segmentation to design a set of structure-specific classifiers. Each individual structure-specific classifier is trained to distinguish benign lesions from melanomas just paying attention to one type of dermoscopic structure. Then, the outputs of the individual classifiers are combined by a means of the Bayesian method that, besides the final diagnosis, provide the doctor with additional valuable information, such as the opinions of the individual structure-specific experts and the uncertainty of the diagnosis.

The results in terms of the features selected for the structure-specific classifiers are consistent with the expert insights. Furthermore, regarding the auto-

Javier López-Labraca · Miguel Ángel Fernández-Torres · Iván González-Díaz · Fernando Díaz-de-María
Signal Theory and Communications Department
Universidad Carlos III de Madrid, Legans, 28911, Spain
Tel.: +34-916248805, +34-916248805, +34-916246262, +34-916249170
Fax: +34-916248749
E-mail: jlabraca@tsc.uc3m.es, matorres@tsc.uc3m.es, igonzalez@tsc.uc3m.es, fdiaz@tsc.uc3m.es

Ángel Pizarro
Clínica Dermatológica Internacional
Madrid, Spain. E-mail: apizarro@ricardoruiz.es

matic melanoma diagnosis problem, the proposed method has been assessed on two different datasets, and the experimental results revealed that the proposed system clearly outperforms other methods in two datasets and compares well with the official submisions of the ISBI 2016 challenge on melanoma detection. Moreover, the system performance is equivalent to that of a well-known dermoscopy expert and its combination with the human diagnosis surpasses the human performance.

Keywords Melanoma Diagnosis · Computer-Aided Diagnosis · Enriched Diagnosis · Dermoscopic Structures · Bayesian Fusion

1 Introduction

Many CAD systems have emerged during the last decades in several medical fields. Illustrative examples can be found in [4], where mammography images are analyzed for automatic tumor detection, in [43], where the authors tackle the Alzheimer diagnosis, or in [10], where visualization techniques are developed to help clinicians in their understanding of the growth of tissues.

Under such a variety of systems, several of them have been proposed for the early diagnosis of melanoma due to various reasons. Melanoma is one of the most aggressive forms of skin cancer, since, as suggested by the latest clinical evidence, once it reaches an advanced stage, it is incurable [22] with the current medicine techniques, and the treatment that the patient would receive at that point (such as chemotherapy or surgery) would only be palliative. In contrast, the early removal of melanoma has shown to completely cure the disease and avoid metastasis [40]. Furthermore, as melanocytes are one of the very few cells that naturally exhibit color, melanocytic lessions are visible to the eye (clinical images) or using portable and specialized acquisition instruments as dermatoscopes (dermoscopic images), which can improve the diagnosis accuracy of clinicians by 5-30% [30].

Therefore, there is an increasing interest on the development of automatic systems helping clinicians to detect this form of cancer in its early stages. However, despite of the many research efforts devoted to this matter, these systems are not yet incorporated to the everyday clinical practice. One of the underlying reasons for such a limitation in their use lies in the fact that most of the systems simply provide a parallel diagnosis to the clinicians, which actually does not help them in practice. These systems usually take advantage of powerful techniques in computer vision, such as the Bag of Visual Words framework (BoVW) and its extensions [8][3], or the more recent Convolutional Neural Networks (CNNs) [13][19]. Although very powerful in terms of classification accuracy, most of these systems lack a clear understanding of the main underlying factors and lesion properties that support their final decision.

This paper proposes a CAD system for melanoma diagnosis that aims to go beyond the simple diagnosis, and focuses on providing additional information of special interest for dermatologists. In contrast to previous works, besides the tentative diagnosis, the proposed CAD system provides the doctor

with a thorough analysis of the specific structures present on lesion, on which the doctor can rely to improve his or her decisions. Moreover, based on a fusion system that properly combines the outputs of specific-structure experts, the system provides the individual opinion of each expert and an uncertainty measure of the system diagnosis.

The remainder of this paper is organized as follows. In Section II an overview of the state-of-the-art methods that address CAD systems for melanoma detection is presented, along with the distinctive characteristics of our proposal. Section III provides a detailed description of the proposed method. In Section IV, the experimental results supporting the proposal are presented and discussed. Finally, Section V summarizes conclusions and outlines future lines of research.

2 Related Work

Two of the most active tasks in the field of CAD systems using dermoscopic images are textural pattern detection and lesion classification. The former aims to outline the different dermoscopic structures (textural patterns) that are present on a lesion. These structures have been defined by expert dermatologists and play an important role in melanoma diagnosis. The latter relies on the analysis of visual information in dermoscopic images to build an automatic model capable of discriminating between benign lesions and incipient melanomas. The intended output of such a model is an automatic diagnosis which assists the clinicians in their decisions.

2.1 Detection of dermoscopic structures

Concerning the detection of dermoscopic structures, dermoscopic structures are the visual result of the actual arrangement of melanin, blood vessels and scar-like tissue [6]. The complete set of structures commonly considered was defined in the *pattern analysis method* for melanoma diagnosis [31], which has been widely adopted by specialists due to its accurate results. Other pattern-based analysis algorithms are simplifications of this one, considering a reduced set of textures or dermoscopic structures which are easier to identify.

Following this approach, many works in the field aim to address the problem of identifying a subset of textures which, in turn, are considered by a classification algorithm to provide a final diagnosis of a clinical case. Quite often, the goal is not only to identify these structures but also to distinguish between their typical and atypical versions. Within this family of methods, generative model-based texture detection assumes that each structure follows a probabilistic mathematical model. The work by Acha et *al* [34] considers that dermoscopic structures can be properly modeled by a Markov Random Field (MRF) with an underlying multivariate normal distribution which is learned during a training phase. A similar idea can be found in [35], where each structure is modeled by means of a Gaussian Mixture Model (GMM). During test, a

new mixture model is inferred and compared with those learned in the training phase, so that each region in the lesion can be classified.

In contrast, discriminative methods aim to find an optimal boundary to separate the training observations into compact groups of the same class. The division of lesions into small regions is a very common step in this case. These regions can arise from a grid division or a low-level oversegmentation. Then, each region is characterized by a vector of features, which are taken from classical approaches to texture recognition. Tanaka et *al* [36] proposed to obtain statistics from the intensity histogram, the Gray Level Co-occurrence matrix (GLCM) and the Run-Length matrix. This approach may lead to a large set of features which can be reduced by means of feature selection algorithms or PCA. Another common approach is based on the use of textons [24] as information units describing complex textural patterns. An example in the field of melanoma screening can be found in [33], where each lesion region is described as a set of filter responses that are clustered to build texton frequency histograms. Other works focus on detecting a particular structure and, additionally, determining if it is typical or atypical. A good example can be found in [28], where Mirzaalian et *al.* identify streaks using features based on the output of a family of tubularness filters and determine if they are atypical. Alternatively, in [7] one can find a survey of methods focused on the reticular structure, covering many image processing techniques for recognition and discrimination between typical and atypical versions of this structure.

2.2 Classification of pigmented skin lesions

The task of classifying pigmented skin lesions is approached as a binary classification problem, where the goal is to determine whether the lesion is benign or malignant. These works usually employ discriminative methods based visual features from the dermoscopic image to provide a tentative diagnosis. Feature extraction varies from the classical machine learning-based approach using general-purpose descriptors (e.g. texture filterbanks [37]) to the problem-dependent knowledge-based features. The first approach relies on a large and problem-independent feature set, which is then pruned using feature selection techniques [1]. On the other side, knowledge-based approaches focus on a reduced set of features derived from expert knowledge and, besides improving the system performance, improve the interpretability of its results [14].

Some works put more emphasis on color information. Barata et *al* [9] proved that estimating the color of the light source for normalization purposes leads to more robust color descriptors. A GMM is trained for each possible dermoscopic color to classify the lesions according to the colors present in them. A task-specific colorspace was introduced in [25], where melanin and hemoglobin are considered the two independent substances that define the colors in pigmented skin lesions and, consequently, two chromatic components are extracted through Independent Component Analysis (ICA). Color histogram statistics and GLCM Haralick features are extracted to train a system for sus-

picious lesion detection. In [2], a large set of texture and color features are obtained and an ensemble method is proposed to make the learning problem more easily scalable. In parcitular, each weak classifier relies on only one feature and the one with highest score is selected on each iteration. Zare et *al* [42] propose a similar approach, in which color features are derived from 8 different color spaces and texture features are computed from the GLCM and a tree-structured wavelet analysis. Different classifiers and feature selection methods are tested and analyzed.

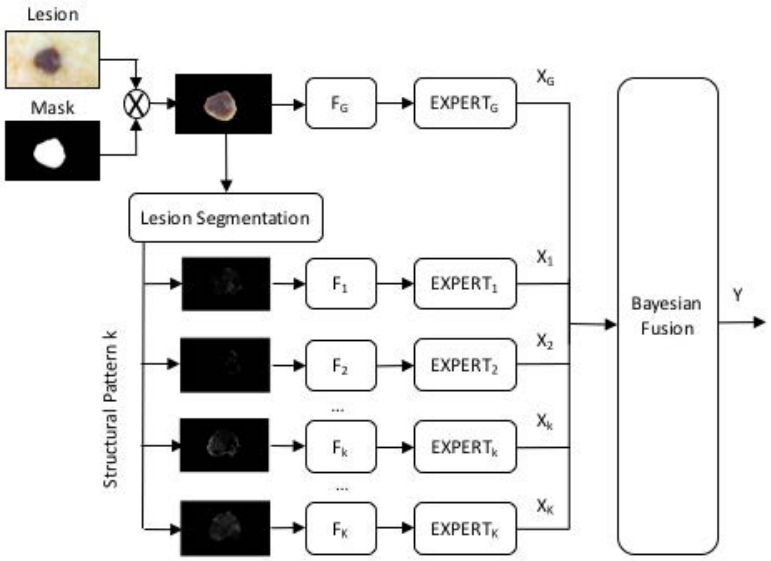
In contrast to the previous approaches, Zortea et *al* [44] proposed a reduced set of interpretable features based on expert knowledge. They model the two normal axes of symmetry and border sharpness from the ABCD rule [29] and train a supervised palette approach based on dermoscopic colors. With simple classifiers (LDA, QDA, decision tree), they obtain similar performance to that of the clinician and better ratio of excisions. Moreover, other works make a complete implementation of diagnosis algorithms followed by dermatologists in their common clinical practice. Here, the classification algorithm is replaced by the punctuation procedure of the corresponding rule. The ABCD rule of dermoscopy is automated in [18] and combined with structure recognition in an attempt to detect suspicious lesions. A set of discriminative and model-based approaches are presented in [23] which provide detectors of the structures that are considered in the 7-point checklist method.

In this paper, we propose a novel approach that relies on the detection of dermoscopic structures to provide the doctor with an enriched tentative diagnosis. Specifically, we focus on a soft structural pattern-based segmentation of the lesion to design a set of specific-structure lesion classifiers, each one paying attention to each structural pattern present on the lesion. On top of this set of specific-structure lesion classifiers, we have developed a fusion scheme which provides a global tentative diagnosis along with additional information aiming to explain its decision (opinions of individual specific-structure experts and a measure of the uncertainty of the decision). In doing so, the proposed system provides significant added value either to improve decisions of an expert doctor or to teach novel doctors.

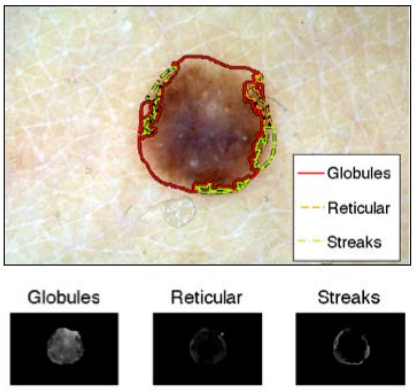
3 The proposed CAD system

In this paper, we propose a Computer-Aided Diagnosis (CAD) system for melanoma detection based on dermoscopic images. Our system, depicted in Figure 1, is hierarchically organized into three main parts:

1. A soft segmentation of the lesion into dermoscopic structures. As explained later, we propose to use soft segmentation maps instead of hard segmentations.
2. A set of individual diagnosis systems, each one of them has been designed for a specific dermoscopic structure. In the following, we will refer to these individual diagnosis systems as *experts*, as they are designed to attain proficiency on the lesion diagnosis by considering just one particular structure



(a)



(b)

Fig. 1: a) Block diagram of the proposed CAD system for melanoma detection. b) Example of lesion segmented into dermoscopic structures. Three structures have been detected in this case: *globules, reticular structure and streaks*, which correspond to three of the experts considered by our approach.

of interest. To build these experts we have proposed a broad set of clinical-based features modeling typical indicators used by dermatologists in their diagnosis. Moreover, we have defined a method to select those particular

features which are more relevant to characterize each of the dermoscopic structures.
3. A Bayesian model which fuses the outputs of the individual experts to provide a final suggestion about the malignancy of a lesion.

It should be noted that the global segmentation of the lesion has been manually defined and is provided to the system as a binary mask together with each skin lesion image.

In the following sections, we will first introduce the set of considered dermoscopic structures and then describe each of the building blocks in our system.

3.1 Considered dermoscopic structures

For each clinical case, our system segments the skin lesion region into dermoscopic structures. It should be noticed that not all patterns appear in every case, thus our system should be able to make a decision based only on those structures detected by the segmentation algorithm. According to the suggestions made by specialists, we have considered eight structures:

1.- *Dots, globules* and *Cobblestone pattern*: although different, they have been fused into one for the purpose of the system development due to their visual similarities. These patterns consist of a certain number of round or oval elements, variously sized, with shades that can be brown and gray-black. In the case of cobblestone structures, they are usually larger, more densely grouped and somewhat angulated. Both are essentially melanocytes nests at the dermoepidermal junction, and they are considered lesion growth indicators when found near the lesion boundary. While an even distribution of them with regular size and shape is associated with benignity, various sizes and shapes, or irregular or localized distribution occur in melanoma.

2.- *Reticular pattern and pigmented networks*: pigmented networks cover most parts of certain lesions. They look as grids of thin brown lines over a light brown background and are quite common in melanocytic lesions. If globally distributed, this structure is related to benign lesions. However, variations in size and form are indicative of malignancy.

3.- *Homogeneous areas*: these areas are diffuse, with brown, grey-black, grey-blue or reddish-black shade, where there is no other local feature that can be recognized. A globally distributed pattern of bluish hue is the hallmark of the blue nevus. With other shades, it may be present in several types of lesions, such as Clark-nevi, dermal nevi or nodular and metastatic melanomas.

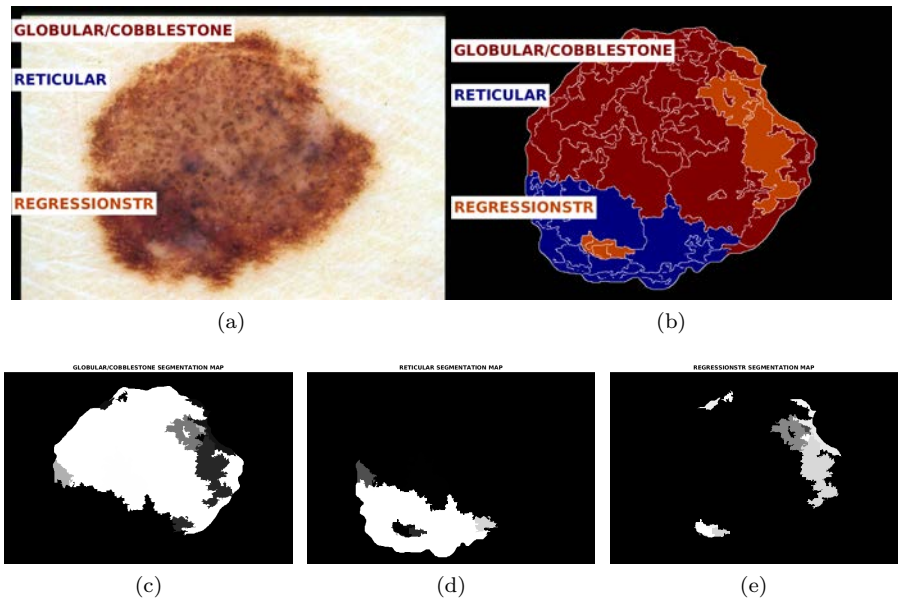
4.- *Regression*: these structures are generally well-defined white and/or blue areas that appear when the immune system has attacked the lesion. White areas resemble a superficial scar, and blue areas may appear as diffuse blue-gray areas or peppering, which is an aggregation of blue-grey dots.

5.- *Blue-white veil*: a region of grey-blue to whitish-blue shade blurred pigmentation, correlated with the presence of heavily pigmented melanocytes and/or melanophages in the dermis in combination with thickened epidermis. It is highly indicative of melanoma.

6.- *Streaks*: are black or light to dark brown longish structures of variable thickness, not clearly combined with pigmented networks, and easily observed when located at the periphery of the lesion. In general, they tend to converge to the center of the lesion. An even, radial distribution of the streaks around the border of the lesion is characteristic of Reed nevus. However, an asymmetric or localized distribution of streaks suggests malignancy.

7.- *Vascular structures*: they are areas with aberrant blood vessels induced by the lesion itself. Depending on their shape, they may be a clear sign of malignancy. While abundant and prominent comma vessels often exist in dermal nevi, some other vascular patterns, such as arborizing, hairpin or linear irregular ones, are more frequent in melanomas.

8.- *Unspecific pattern*: we group in this category those parts of the lesion that cannot be assigned to any of the previous structures. No direct diagnosis implication can be inferred from it. Nevertheless, it is more often related to melanoma, or at least it suggests that the lesion must be carefully explored.



(a) (b)

(c) (d) (e)

Fig. 2: a) Weakly-annotated image for the segmentation. (b) Segmentation provided with our approach. White boundaries represent previous segmentation in superpixels used by the algorithm in [15]. (c-e) Soft segmentation maps $P_k = \{p_k(\mathbf{x})\}$ for dermoscopic structures dots/globular/cobblestone, reticular, and regression, respectively. Lighter pixels indicate higher probabilities of belonging tot he structural pattern.

3.2 Soft segmentation maps for feature computation

One of the key ideas of this paper is to segment the dermoscopic structures present in the image to extract a specifically tailored set features for each structural pattern. Moreover, we have preferred to work with soft maps containing probabilities rather than with a hard segmentation in which each pixel is only assigned to the most probable structural pattern. Therefore, we first generate these soft maps and then we extract the features (for each pattern) weighted according to these maps.

For the generation of the soft segmentation maps, we have made used of a modified version of the segmentation approach proposed in [15], which combines Latent Topic Models (LTM) and a Kernel Logistic Regression (KLR)-based appearance models to generate category-based segmentations under both supervised and unsupervised scenarios. The application of this algorithm to our current problem would require a training dataset in which each image had a corresponding ground truth segmentation. This kind of (strong) annotation is often hard to obtain as it requires a huge effort from the dermatologists to outline the segmentations. Alternatively, providing weak image-level annotations indicating only the presence of the different dermoscopic structures is much easier for dermatologists and therefore becomes a more feasible approach. Indeed, one of the considered datasets in this work already contained these weak annotations and the other was annotated by an specialist in dermoscopy. Thus, our training datasets were finally annotated using 3 image-level category labels for each structural pattern: 0, if the structure is not present; 1, if it is locally present; and 2, if the structural pattern takes up most of the lesion area and can be then considered as global.

The resulting weakly-supervised scenario requires to adapt the original approach to work with image-level annotations. For the sake of simplicity, we only introduce here the notation required for integrating the resulting soft segmentation maps into subsequent stages. The interested reader can find an in-depth analysis of the algorithm as well as the detailed formulation in the original paper [15].

Given an input image $I(\mathbf{x})$ and the mask $M(\mathbf{x})$ that defines the region of the lesion over pixels $\mathbf{x}$, the weakly-supervised segmentation algorithm provides a set of K soft segmentation maps $P_k = \{p_k(\mathbf{x})\}$, where k is the index of the structural pattern $k = 1...K$ (as described in sec. 3.1), and $p_k(\mathbf{x})$ stands for the probability that the pixel $\mathbf{x}$ belongs to the structural pattern k. Moreover, we impose a unit simplex constraint so that $\sum_k p_k(\mathbf{x}) = 1$ for each pixel in the lesion.

In our approach for melanoma automatic diagnosis, low-level features are first locally computed at each pixel location, and then aggregated at the image level to produce lesion-level descriptors. When computing a feature for a particular structural pattern k, it is intuitive that the pixels belonging to that structure should have a higher impact on the final descriptor than those on the rest of the lesion. Consequently, we have incorporated the segmentation map $p_k(\mathbf{x})$ into the feature computation process. In most cases, the feature

aggregation is a simple linear combination of pixel outputs; in particular, a feature F_k^d computed over image I^d and designed for the structural pattern k is obtained as follows:

$$F_k^d = \sum_i^N p_k(x_i) f_k(x_i) \tag{1}$$

where N is the number of pixels within the lesion area, and $f_k(x_i)$ is the value of the feature evaluated in the pixel x_i. Examples of $f_k(x_i)$ can be the energy of a filtered image in x_i, the values of the color components in that pixel, etc. As we will see in the following sections, although linear feature aggregation is the most common approach in our system, for some features like shape-related ones, we apply thresholding techniques over $p_k(x_i)$ to generate a binary mask over which computing the descriptors.

3.3 Feature extraction

This section describes the feature extraction process followed in our approach. We first introduce a new coordinate system that represents more appropriately the geometry of the lesions. Then, we review the families of filter-banks on which we have built the extraction of texture features. Subsequently, we describe the features, which we have divided into two groups: generic and specific features, where the last ones have been specifically tailored to detect or measure properties of a given structural pattern. Finally, we explain the feature selection process followed for each expert.

3.3.1 Normalized polar coordinates

Based on recommendations of dermatologists, and due to the lack of knowledge about the actual lesion sizes, we have designed a *normalized polar coordinate system* that allows comparing equivalent spatial locations of different-size lesions. For each lesion d, defined by the pair of image I^d and mask M^d, we first model the shape of the lesion as an ellipse and compute the centroid $\mathbf{x}_0$, the lengths a, b of the major and minor axes, respectively, and the orientation angle θ. Based on these values, we define the scaling S and rotation Q matrices as:

$$S = \begin{pmatrix} 1 & 0 \\ 0 & a/b \end{pmatrix}; \quad Q = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{2}$$

Then, we can compute an affine 3x3 matrix T that transforms the input ellipse into a normalized (unit radius) circle, centered at (0,0), as:

$$T = \begin{pmatrix} C & d \\ 0 & 1 \end{pmatrix} \tag{3}$$

where the block matrix C is $C = QSQ'$ and the vector $d = (I - C)\mathbf{x}_0$. This transformation will be thoroughly used along the paper as it allows us to

first obtain the normalized Cartesian coordinates $\mathbf{x}_n = T\mathbf{x}$ on the transformed unit circle and, then, to compute the normalized polar coordinates $(\rho(\mathbf{x}_n), \theta(\mathbf{x}_n)), \rho \in [0,1], \theta \in [0, 2\pi]$ associated with each pixel in the lesion. This new coordinate system will be used during the computation of various features in our model, such as those related to lesion symmetry. It is noteworthy that we do not explicitly transform the input image as it would lead to artifacts and distorted visual representations.

3.3.2 Enhancing texture through filtering

Texture features are not extracted directly from images. In order to make them more robust to illumination changes and transformations, images are preprocessed using a set of filters developed to detect specific texture properties in dermoscopic images. Specifically, we have used two families of filter-banks: a) multi-scale Laplacian of Gaussian (LoG) filter-bank, and b) multi-orientation multi-scale Gabor filter-bank.

The multiscale LoG filter-bank works as a common texture enhancer at its smallest scales and as a blob detector at its middle and large scales (see Figure 3). An initial set of filtered images is obtained using LoG filters at different scales. A σ^2 normalization is applied to every output to make them comparable in the scale space. Once the output of each filter is computed, non-maxima suppression is applied across scales to find the scale with highest response at pixel level, i.e., for each scale only those pixels that provided the strongest response keep its value (while the others are set to zero). Additionally, a multi-scale summary image is obtained by fusing the outputs at different scales after the non-maxima suppression.

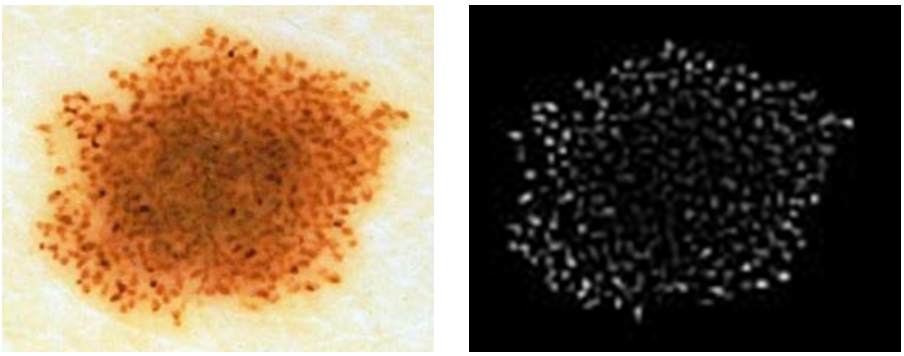
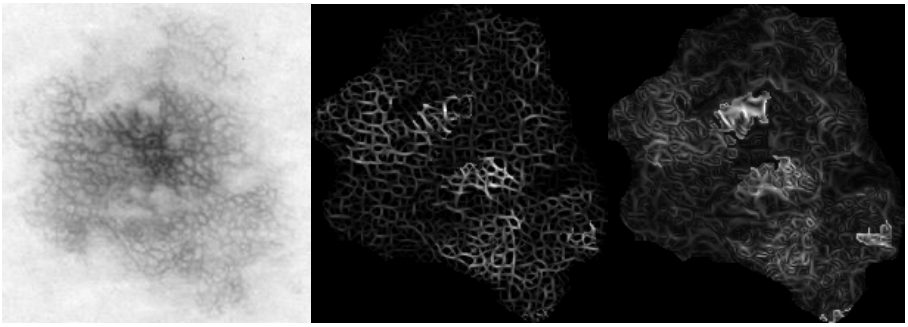


Fig. 3: Multiscale LoG filter works as a blob detector for globules.

The multi-orientation multi-scale Gabor filter-bank works as a border detector. The family of Gabor filters [41] offers two different base functions to detect specific types of transitions (even and odd). Both of them are used in parallel to get complementary descriptions of border presence in a lesion. For

each family, a set of filtered images is obtained varying orientation and scale. Then, a non-maxima suppression process is applied in the space of directions and scales. As a result, each filtered image for a specific orientation-scale pair contains non-zero outputs only where it was maximum. After non-maxima suppression we integrate over all orientations to get a rotation-invariant set of filtered images, so that we reduce the original direction-scale space to the scale space as with LoG. Again, a composite image summarizing the maximum responses in the scale space (see Figure 4) is computed.

In summary, the output of the texture-enhancing filtering stage for a given image I^d and a given filter family h_s is a set of output images o_s , computed using the cross-correlation $o_s = h_s \star I^d$.



a) b) c)

Fig. 4: Gabor filters highlight the lines of the grid in a lesion with reticular structure. a) Original Image; b) even filters; and c) odd filters.

3.3.3 Generic features

The generic features are intended to describe the global aspect of a lesion and properties that are common to various of the considered structures. The following paragraphs characterize generic features according to four different groups: geometry, symmetry, color and texture.

Geometry We propose two types of features related to geometric aspects of the lesion: first, a simple feature describing the *area* of a structure relative to that of the lesion; and second, more complex features that describe the *spatial distribution* of dermoscopic structures. Table 1 summarizes the set of features explained below.

For that feature related to the area, as it will be explained later in Section 3.4, the region of a specific structure k is outlined by means of a thresholding operation on the corresponding probability $p_k(x)$. Subsequently, the resulting region area is normalized with respect to the area of the lesion.

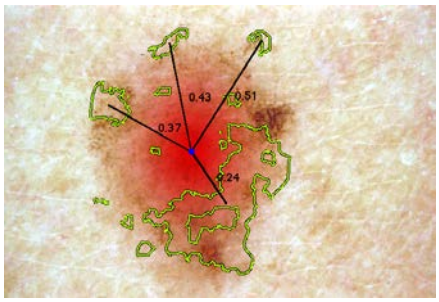


Fig. 5: Radial distance and spread of the Globules, measured as the weighted mean and variance of the distance of the structure pixels to the center of the lesion, respectively. Distances from the center of the lesion to the centroids of each of the regions detected as Globules in the example are indicated.

The second group of geometric features are inspired by the fact that some structures can be indicators of benignity or malignancy depending on their relative location in the lesion. In dermoscopy literature, this is the case of peripheral globules, homogeneous blobs or reticular structures. Here, the structure probability $p_k(\mathbf{x}_i)$ is used as soft indicator of the presence of the structure and it allows us to estimate its relative location with regard to the center of the lesion, from which we are able to compute statistics about its distance to the center of the lesion and its angular distribution. For the first aspect, the distance of each pixel to the center of the lesion is computed, as illustrated in Figure 5, where the blue dot represents the center. Then, weighted mean and standard deviation (using $p_k(\mathbf{x}_i)$) are obtained, modeling the radial mean distance and the radial spread of the local structure, respectively. For the second aspect, the angle that each pixel forms with respect to the horizontal is computed and weighted by $p_k(\mathbf{x}_i)$. The circular standard deviation of the weighted angles describes the angular spread of the texture over the lesion, denoting if the structure is located at a specific sector or widespread over the lesion.

Table 1: Geometry features measure spatial aspects of the structure with respect to the lesion, considering area and spatial distribution.

Group	Name	Description
Area	Relative area	Ratio area of the detected structure / area of the lesion
Spatial distribution	MeanDist	Mean distance to the center of the lesion
	StdDist	Mean std of distances to the center of the lesion
	MeanAng	Mean angular position of the structure
	StdAng	Angular spread of the structure

Symmetry features Melanomas tend to grow differently on each direction, becoming more asymmetric than benign lesions. This is why symmetry is present

Fig. 6: Polar grid division for the computation of the main axis of symmetry (bold red) and its complementary (red). Each cell in the polar grid is then represented by its average color (right image).

in a variety of diagnosis algorithms, such as the ABCD rule of dermoscopy [29]. It includes shape and color as different aspects to evaluate symmetry in pigmented skin lesions. The symmetry rule requires finding the axis of maximum symmetry according to any of those two criteria, and its perpendicular. In doing so, the lesion is labeled as symmetric in one or two axes, or as completely asymmetric.

In this work, we have designed a real-valued version of this symmetry rule which is computed as follows: the lesion is first divided into a discrete grid with sectorial and radial partitions (see Figure 6) using the normalized polar coordinates described in Section 3.3.1. Then, for each given angle, the lesion is divided into two halves and each cell of the grid is compared with its symmetric with respect to the axis. In order to get the actual similarity between each pair of cells, we have used pixel color distances in the CIE Lab colorspace, as it is close to human perception. Hence, the symmetry feature associated with that angle is the average similarity between all the corresponding pairs of sectors in the partition. Finally, the angle with the maximum average similarity and its orthogonal are selected as symmetry axes and the average similarity values are considered as the features. These features are also computed for every structural pattern, by using the soft weights described in Section 3.2.

Color features Color is one of the essential clues for the diagnosis of pigmented skin lesions [31]. Dermoscopic images can show basically the following colors: black, grey, dark brown, light brown, blue, white and red. The first five are related to the depth of melanin at the epidermis, while the last two are indicators of scar-like tissue and blood vessels, respectively. Thus, the number of colors on the lesion, so as their spatial distribution, give information about the distribution of melanin through the nevus. To model different aspects of color information, we use a set of color descriptors:

- The color-naming descriptor developed by Benavente, Vanrell and Baldrich [11] provides a histogram of presence of 11 common colors.
- The discriminative hue descriptor by Khan et *al* [20] obtains a set of photometric invariant color groups.
- The opponent color descriptor by van de Weijer and Schmid [38] brings a color histogram based on an opponent representation of color components.
- The extended color-naming descriptor proposed by van der Weijer et *al* [39] offers a color histogram considering 24 bins related with common colors.

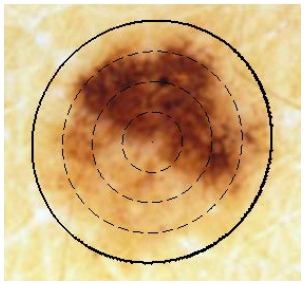
Each of the color histograms provided by these descriptors are obtained in multiple points of the lesion using a dense grid. Then the mean and variance of the histograms computed at all positions are taken as the color descriptors over the lesion. As in the case of symmetry features, the color features are also computed for every structural pattern using the soft weights. Table 2 shows the considered combinations between color descriptor and aggregation measure, each of them is treated as a different feature.

Table 2: Color features obtained from 4 different color descriptors. The features act as global color representations of the lesion.

Color descriptor	Aggregation
Color naming	Mean
Color naming	Standard deviation
Discriminative	Mean
Discriminative	Standard deviation
Opponent	Mean
Opponent	Standard deviation
Extended color naming	Mean
Extended color naming	Standard deviation

Texture features These generic features aim to describe the visual appearance of textures along a lesion with the purpose of modeling the clinical insights concerning dermoscopic structures. Using the filter-banks introduced in Section 3.3.2, we first compute the corresponding sets of filtered images for the lesion. For these sets of outputs, two groups of features are computed: a) features relying on ring- and circular sector-based partition analysis; and b) features that evaluate the texture regularity working on the energy histograms at different scales.

The *ring- and circular sector-based features* are based on a grid division of the lesion using the normalized polar coordinates. The number of radial and sectorial divisions was determined through a cross-validation process, which gave 4 radial and 12 angular partitions as optimal values. Each considered feature is completely defined by three elements: 1) the division of the lesion (radial, sectorial or both, as illustrated in Figure 7; 2) the statistical metric computed at each cell in the grid (mean, variance or histogram of energies



(a) Rings (b) Sectors (c) Cells

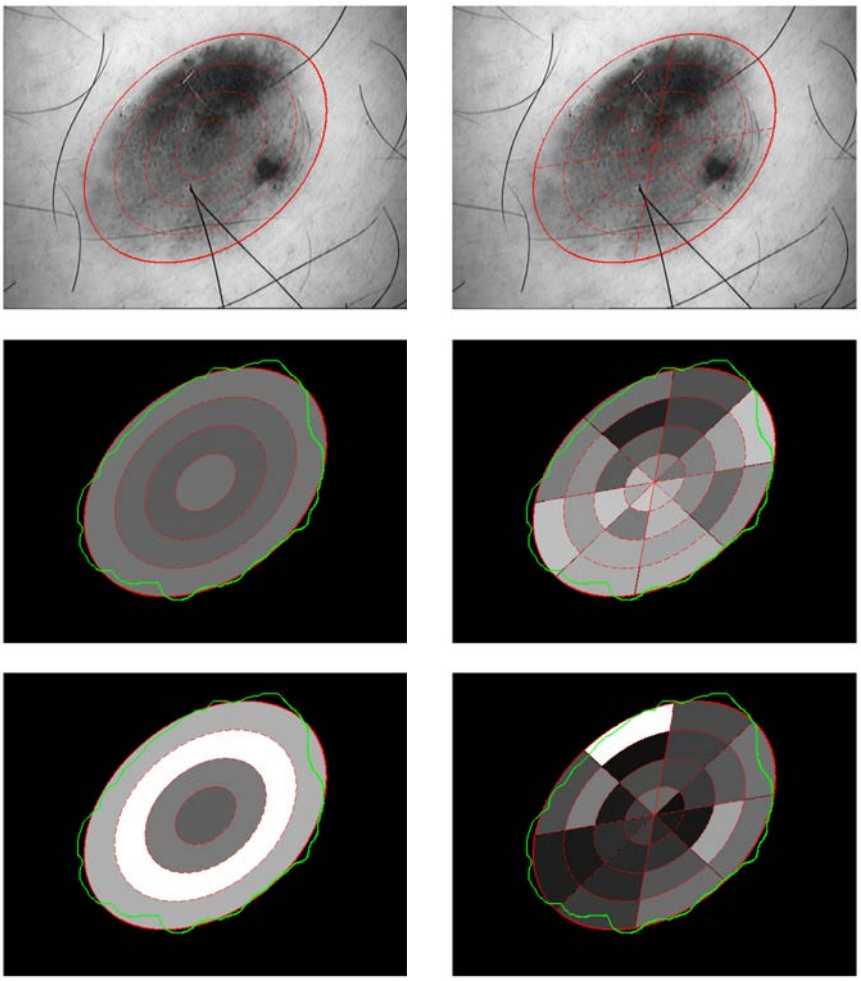
Fig. 7: Different skin lesion division patterns used to compute generic features.

in the scale space); and 3) the aggregation method used to compute a final feature from the individual measures in each cell (mean, variance, simple concatenation or symmetry computation). In Table 3 we provide a compact description of the different texture features considered in our approach. In addition, Figure 8 shows two examples of these texture features.

Table 3: Texture features in rings and circular sectors

Name	Division	Cell Metric	Aggregation
Var of mean energy in rings	Rings	Mean	Var
Var of variance energy in rings	Rings	Var	Var
Var of mean energy in sectors	Sectors	Mean	Var
Var of mean variance in sectors	Sectors	Var	Var
Var of mean energy in cells	Both	Mean	Var
Var of variance energy in cells	Both	Var	Var
Var of mean in rings	Rings	Mean	Var
Var of variance in rings	Rings	Var	Var
Var of mean in sectors	Sectors	Mean	Var
Var of mean in sectors	Sectors	Var	Var
Var of mean in cells	Both	Mean	Var
Var of variance in cells	Both	Var	Var
Mean of rings energy	Rings	Mean	Concat
Var of rings energy	Rings	Var	Concat
Mean of two axes of symmetry	Both	Mean	Sym
Var of two axes of symmetry	Both	Var	Sym
Hists of two axes of symmetry	Both	Hist	Sym

A second group of texture features evaluates the *texture regularity* and is based on the analysis of histograms. In dermoscopy, a structure is considered typical when it exhibits a regular appearance. This aspect is intended to be modeled by this group of features, which first run a filter-bank and then compute the histogram of energies in the scale space. The presence of high energies at different scales is an indicator of non-regularity of the elements that form a structure, whereas regularity is paired with a dominant scale and, therefore, a



a) b)

Fig. 8: Texture feature extraction for an atypical lesion. First, a grid division of the lesion is obtained using the normalized polar coordinates, and different statistics are computed at each cell in the grid. Finally, an aggregation method is used to compute a final feature from the individual measures in each cell. a) Variance ratio of mean and variance energy in rings. b) Variance ratio of mean and variance energy in cells.

prominent peak in the histogram. Given an image I^d, a structural pattern k being analyzed, the topic weights $p_k(\mathbf{x}_i)$ (sec. 3.2), and the set of outputs o_s of a filter family, the histogram $H_k(s)$ can be computed as follows:

$$H_k(s) = \sum_{i=1}^{N} p_k(\mathbf{x}_i) o_s^2(\mathbf{x}_i) \tag{4}$$

where s stands for a discrete scale. However, as a texture not only provides a high response to its best matching scale but also to its adjacent scales, the bars of the histogram are not completely independent. To correct this undesired effect, an overlapping-based correction is applied to the histogram. The overlapping factor $O(s_i, s_j)$ between two scales s_i and s_j is first computed and, then, the histogram is corrected as follows:

$$\hat{H}_k(s_i) = H_k(s_i) - \sum_{s_j \neq s_i} H_k(s_j) O(s_i, s_j) \tag{5}$$

In order to compute the overlapping factor $O(s_i, s_j)$, we have approximated the spatial extent of the filters using 2-D Gaussian distributions, and computed the Bhattacharyya coefficient as:

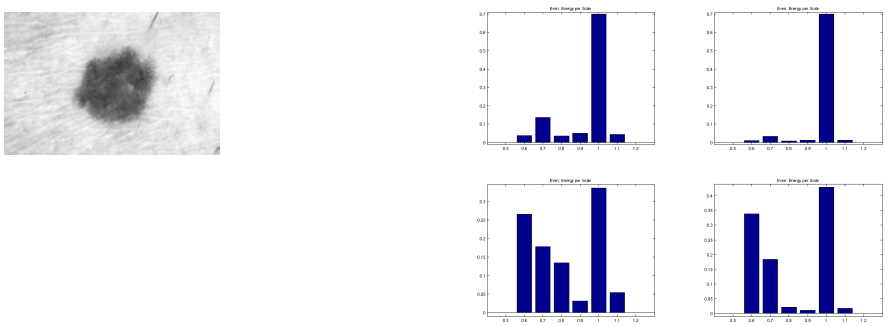
$$O(s_i, s_j) = 2\sqrt{\frac{C_{ij}^2}{1 + C_{ij}^2}} \tag{6}$$

where $C = \frac{\sigma_{s_i}}{\sigma_{s_j}}$ is the ratio between the standard deviations of the filters, and we have considered zero means as the filters are always centered on the pixel of interest.

Once the histogram has been corrected, the texture features are extracted from it. Table 4 briefly describes the set of features we have designed to evaluate the texture regularity, and Figure 9 provides two examples.

Table 4: Features measuring texture regularity, calculated from the histogram of energies as a function of the scale.

Name	Description
Entropy	Entropy of the histogram
Peaks50Perc	Number of peaks over the 50th percentile
Peaks75Perc	Number of peaks over the 75th percentile
SigmaMax	Scale with the highest global energy
EnergyOutside	Energy out of the dominant scale and its two neighbors
SigmaMean	Expectation of the histogram of energies
SigmaStd	Standard deviation of the histogram of energies



a) b) c) d)

Fig. 9: Top row is an example of typical reticular structure, while bottom one shows an atypical network. a) Original image. b) Multi-scale image that summarizes the maximum responses of the Even Gabor family of filters in the scale space. c) Histogram of energies at each scale. d) Corrected Histogram.

3.3.4 Structure-specific features

Besides the aforementioned generic features, we have considered specific attributes of the structures that are important for the dermatologists in their decisions. We call them structure-specific features since each one has been designed for a particular structure or pattern. We describe them in the following paragraphs. As in the previous cases, the soft segmentations $p_k(\mathbf{x})$ (see Section 3.2) indicate the region of the lesion where the presence of each structure is more likely.

- *Dots, globules* and *Cobblestone pattern*: since peripheral globules are related to lesion growth, the multi-scale LoG filter output is used to compute a ratio between the energy of globules in the central part of the lesion and the energy of globules in the border.

- *Reticular pattern*: we suggest a specific descriptor to determine its prominence, which is a clinical indicator of malignancy considered by some dermoscopic algorithms, such as the Menzies' method [27] and the Seven Point Checklist [5]. We rely on the outputs of multi-scale LoG filter bank to compute the energy of the light blobs corresponding to the center of the reticular cells and the energy of the dark lines defining the reticular cells. The ratio between the mean energies of these two regions provides a measure of the contrast, which can be used as a detector of prominent reticular texture.

- *Homogeneous areas*: this structure is commonly found in the center of the lesion. As its peripheral presence is less usual, a potentially malignant radial growth[1] of uniform blobs is searched for. The lesion is first divided into rings. The mean amount of the structure at each ring, moving from the center to the border, is then computed and saved into a vector. Next, the first derivative

[1] The word 'growth' is employed here with the meaning 'increase of structure presence as we move out of the center of the lesion'. It is not related to lesion growth.

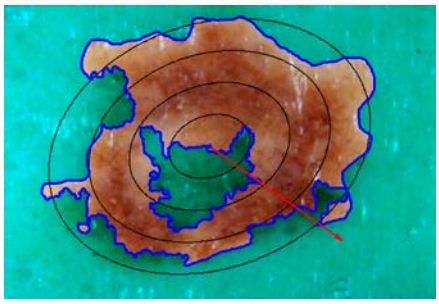


Fig. 10: Radial growth of homogeneous areas. The first derivative along rings is obtained to find out evidences of growth. Its cumulative sum provides a growth indication of the presence of homogeneous structure when going out of the center.

is obtained to find out evidences of growth. Negative values are turned to 0, as they mean lack of growth. The cumulative sum of positive values, if any, provides a growth indication of the presence of homogeneous structure when going out of the center. Figure 10 illustrates the process.

- *Regression structures*: the average luminance ratio between the regression area and the rest of the lesion is considered. The purpose of this feature is to distinguish between white and grey regression.

3.3.5 Feature selection for individual experts

Due to the huge amount of extracted features, it is necessary to determine which ones are more descriptive for each structural pattern. To that end, a forward-feature selection method using linear Support Vector Machines (SVM) [1] is applied that efficiently chooses the minimum group of relevant features that jointly maximize the benign vs. malignant classification performance, measured by the Area Under the Curve (AUC). Hence, for each structure feature set, we perform an N-fold cross-validation procedure, as it is shown in the pseudocode in Algorithm 1. Starting from an empty set of features, we add the feature that mostly improves the AUC at each time, and iterate until AUC reaches its maximum value.

The selected features are then used to train a classifier associated with each structural pattern k, i.e., an *expert* on each structural pattern. The soft outputs $X_k, P_k \in (-\inf, \inf)$ of these experts will be then fused as described in the next section to generate the final diagnosis. In addition to the K considered dermoscopic structures, we have also added a global expert X_G, which is structure-agnostic, and works over the whole image without considering any soft segmentation ($p_G(\mathbf{x}) = 1$). The candidate input features for this expert is the generic set of features described in Section 3.3.3.

Algorithm 1 Forward-feature selection using SVMs

Input: Whole d-dimensional feature set $F = \{f_1, f_2, ..., f_d\}$, number of folds N
Initialization: Empty subset $S_0 = \emptyset$, size of the subset $k = 0$, $AUC_{max} = 0$,
while $\sim isempty(F)$ **do**
  $AUC_{kmax} = 0$
  for $i = 1$ **to** $length(F)$ **do**
    $S_{new} = \{S_k, f_i\}$
    for $n = 1$ **to** N **do**
      Train $model_n$ on remaining $n - 1$ folds.
      Predict n-fold labels using $model_n$.
    end for
    Compute current new subset AUC_{ki}.
    if $AUC_{ki} > AUC_{kmax}$ **then**
      $ind = i$
      $AUC_{kmax} = AUC_{ki}$
    end if
  end for
  if $AUC_{kmax} > AUC_{max}$ **then**
    $S_k = \{S_{k\ 1}, f_{ind}\}$
    $AUC_{max} = AUC_{kmax}$
    Remove f_{ind} from F.
    $k = k + 1$
  else
    $S = S_k$
    break
  end if
end while
Output: Subset of features S, number of features selected k, AUC_{max}.

3.4 Fusing experts' outputs

For each input image in our dataset, we first identify the dermoscopic structures that appear in the lesion using the output of our segmentation module described in Section 3.2. To make this decision, we consider the maximum probability of a pixel belonging to a structure $P_k = \max\{p_k(x)\}$ and compare it with a threshold. Based on some previous experimentation, we have found that a threshold of 0.2 provides reasonable results in our datasets.

Given a particular clinical case, the input to the fusion stage is a feature vector $X = \{X_1, X_2, ..., X_M\}$ containing the soft outputs of the M experts available for the case (the dermoscopic structures that appear in the lesion plus the global expert). In general $M \leq K + 1$, with $K = 8$ in our case. Therefore, we need a flexible approach able to handle missing data of those dermoscopic structures that are not present in the lesion. Our goal is to generate a final diagnosis, Y, a measure of the experts' individual opinions O_m on the system diagnosis, and a measure of the uncertainty U of the decision according to variance of the experts' individual opinions. To this end, we have used a Bayesian fusion scheme which was successfully applied to MRI image classification in [26]. This model assumes that the likelihood of the data provided by the m-expert conditioned to the class j ($j = 0, 1$ for benign and malign lesions, respectively), $p(X_m|Y_j)$, is independent from the remaining

experts n ($\forall n \neq m$). Hence, we can easily factorize the joint likelihood of the experts' opinion $p(X|Y_j)$ and, using Bayes' theorem, model the posterior probability of a class j:

$$p(Y_j|X_1...X_M) \propto p(Y_j) \prod_{m=1}^{M} p(X_m|Y_j). \tag{7}$$

Consequently, the final diagnosis is computed as:

$$j = \arg\max_j p(Y_j|X_1...X_M) \tag{8}$$

and depends on the *prior* probability of the classes $p(Y_j)$ and the likelihood of the inputs $p(X_m|Y_j)$. This model, known as *Independent Likelihood Pool* [12], allows to implicitly handle missing data of those dermoscopic structures that are not present in the lesion, by simply removing them from the factorization.

As the soft outputs of an SVM X_m, with $m = 1, ..., M$, are continuous and unbounded, we have modeled the corresponding likelihoods as Gaussian random variables $p(X_m|Y_j) = N(\mu_{jm}, \sigma_{jm})$, defined by their mean μ_{jm} and standard deviation σ_{jm}. Modeling the experts' outputs as Gaussian distributions, rather than directly using their values, allows us to not only handle missing data, but also to measure the uncertainty over the outputs of the experts.

In addition to the final diagnosis, we can also provide extra valuable information for the dermatologists. First, we can compute *Individual expert opinions* O_m as:

$$O_m = p(Y_j|X_m) \propto p(Y_j)p(X_m|Y_j), \tag{9}$$

being j the finally selected class. An opinion O_m tells us about the degree of agreement of the m-expert output with the global diagnosis. Therefore, these expert opinions lets us to know the dermoscopic structures that become essential (due to their degree of agreement) to the final system decision. As we will discuss in section 4.5, this information might be of great interest during the training of new medical staff.

Moreover, we can measure the *Uncertainty U of the final decision*. This metric is defined as the mean deviation between the individual expert opinions $p(Y_j|X_m)$ with respect to the final output of the fusion stage $p(Y_j|X_1...X_M)$, where j is the selected class:

$$U = \sqrt{\frac{1}{M} \sum_{m=1}^{M} \left(p(Y_j|X_1...X_M) - p(Y_j|X_m)\right)^2} \tag{10}$$

This uncertainty metric provides some insight into the reliability of the system decision.

4 Experiments and Results

4.1 Datasets and experimental protocol

Two databases have been employed in our experiments. The first one has been obtained from the EDRA Interactive Atlas of Dermoscopy [6]. Its images were gathered with the objective of providing a panoramic view of skin lesions diagnosis using dermoscopy and they are frequently used in works about pigmented skin lesion classification. All the images were saved in JPEG format with almost uniform sizes (768x512). They acquisition process does not follow any acquisition protocol. The ground truth labels are determined by histopathological diagnosis. There are no metadata associated with the image files, so both the device and the zoom level used are unknown. Non-melanocytic images and lesions on palms or soles are discarded, because they are out of the scope of the study. This gives a total set of 724 melanocytic lesions, 222 of them are melanomas and the remaining 502 are benign lesions. Every image of this dataset has, as associated elements, a set of weak labels coming from the database and a mask which identifies the region associated with the lesion, which has obtained by manual segmentation.

The second dataset contains images loaned by two institutions: *Clinica Dermatologica Internacional (CDI)* and *Instituto Madrileo de Oncologia (IMO)*, both in Madrid, Spain. This dataset, hereafter referred to as CDI-IMO dataset, consists of images taken during the clinical practice in several hospitals, so that the acquisition equipment, image size and quality are variable. The set of acquisition devices includes *Molemax-II*, *Molemax HD*, *Dermlite Photo*, *Dermlite DL3* and *Heynedermaphot*. All of them work with polarized light, except for the last device. Image sizes vary from 640x480 pixel for *Molemax-II* to 2592x1944 for *Dermlite Photo*. The inclusion of this dataset pursues the goal of testing the proposed system on images that are hard for the specialists themselves. Whereas the examples of lesions included in an atlas of dermoscopy are illustrative for educational purposes in the field of dermoscopy, clinical practice is expected to be more challenging as dermatologists aim to excise melanomas at their earliest stage, where only few and slight evidences are present in the lesion. In this scenario, atypical benign lesions and melanomas may show very similar aspect. This set of images is specifically composed of cases with hard diagnosis difficulty.

For both databases, we have conducted our experiments following the same protocol: the dataset has been split into two sets: 80% of images are used for training and the remaining 20% for test. Each lesion has been first soft segmented into dermoscopic structures and, depending on those patterns that are present, the corresponding *experts* are run over their specific sets of features (see Section). In order to select the most appropriate C parameter for the SVM classifier behind each expert, the training set has been in turn divided following a 5-fold cross validation. Additionally, this cross-validation also allows us to learn optimal distributions on the training data for the Bayesian Fusion. In particular, the validation subset for each of the 5 partitions are concatenated

and used to learn the parameters of the Gaussian distributions, thus avoiding overfitting. Therefore, both the individual experts and the Bayesian fusion are evaluated on the unseen test set.

Due to the limited size of the datasets, and with the purpose of providing reliable results, the whole process has been repeated 100 times, considering different random partitions for training and test sets (but keeping the 80-20% proportions). Finally, the performance of the proposed approach has been evaluated in terms of AUC.

Finally, for the sake of facilitating the comparison with our system over the EDRA (which is commercially available), we provide the EDRA [6] database indexes and the 100 partitions used in the following link[2].

4.2 Analysis of selected features

Using the forward-feature selection method described in Section 4.1, we have selected the most relevant features for each expert. The results are summarized in Table 5, where we refer to the general types of features to provide a more convenient and intelligible perspective. Otherwise, it would be difficult to interpret by the readers since the list of specific features is long and their denomination is quite technical (see Tables 3 and 4, for example). In particular, in Table 5 we refer to the four types of general features described in Section 3.3.3, i.e., geometry, symmetry, color, and texture; and to the pattern-specific features described in Section 3.3.4, which are shown in italics. Furthermore, more detailed information is provided in parenthesis when it is pertinent for the discussion.

In the next paragraphs we discuss the results obtained for each one of the dermoscopic structures, giving an interpretation according to the intuitions of the dermatologists. But, before starting the discussion of specific dermoscopic structures, it is in order to mention that clinical symmetry features turn out to be useful to characterize almost every one, which matches perfectly the intuitions of the dermatologists since melanomas become more asymmetric than benign lesions.

4.2.1 Globules:

Geometric information and, in particular, statistics about the distance to the lesion center are relevant to describe *globules* spread, as illustrated in Figure 5. As dermatologists suggest, the presence of peripheral *globules* is representative of lesion growth. Besides, a ring of regularly distributed peripheral globules is characteristic of many growing melanocytic nevi. In addition, texture regularity measures help to identify the presence of globules at various scales, which might be indicative of melanoma.

[2] We are currently preparing the dataset, which will be available in the upcoming weeks.

Table 5: Most relevant feature types selected for experts. Features are ordered by their relevance. Pattern-specific features in italics (described in Section 3.3.4). See Tables 1-3 to look over the features included in each category.

Globules/Cobblestone	**Blue-White Veil**	**Reticular Structure**
Symmetry	Color	Symmetry
Geometry (spatial distribution)	Symmetry	Geometry (spatial distribution)
Texture regularity	Geometry (area)	Texture regularity
	Geometry (spatial distribution)	

Homogeneous Areas	**Regression Areas**	**Streaks Structure**
Color	Symmetry Color	Symmetry
Geometry (spatial distribution)	Texture regularity	Texture regularity
Radial growth	Geometry (spatial distribution)	
Luminance Ratio		

Vascular Structure	**Unspecific Pattern**
Texture regularity	Geometry
Geometry (spatial distribution)	Texture regularity
	Symmetry

4.2.2 Blue-white veil:

Blue-white veil is properly described by color representations [20,38]. Although the presence of this structure is almost always an indicator of melanoma, geometry features such as its relative size with respect to the lesion area, and its spatial distribution are also helpful for the corresponding expert.

4.2.3 Reticular structures:

As it can be seen in Figure 4, the reticular structure is better represented using Gabor even filters outputs. Hence, metrics about texture regularity allow the expert to distinguish between typical (thin line, moderate brown pigmentation, regular holes) and atypical networks (thick line, intense dark pigmentation, irregular holes). Figure 9 illustrates this difference with two examples of typical and atypical reticular patterns, respectively, and their corresponding histograms of texture scales. It is easy to notice how the atypical case produces a more complex histogram. We also have found that the spatial location of the pattern is also relevant for the diagnosis.

4.2.4 Homogeneous areas:

Most of the color descriptors are advantageous to describe homogeneous areas. Spatial distribution features, i.e., statistical measures of the dispersion with respect to the center of the lesion and of the angular distributions, are also of special interest for these structures. Furthermore, it is worth mentioning the selection of a feature specifically designed for this expert: the *radial growth* (see Figure 10).

4.2.5 Regression areas:

As for homogeneous areas, color and spatial distribution features are also useful to describe regression areas. Likewise, it is also worth mentioning the selection of a feature specifically designed for detecting regression areas: the *luminance ratio*. Moreover, texture features have also shown to be relevant to detect grey regression, which normally appears as a peppering structure that is more malignant when is large.

4.2.6 Streaks:

They are well defined by analyzing their texture regularity. It was somewhat surprising the lack of geometry features as it is known that the irregular and partial presence of streaks in the lesion borders are key clues for the diagnosis. In our opinion, symmetry features are already encoding this property in this case.

4.2.7 Vascular structures:

The joint analysis of textures and spatial distribution becomes very useful. In particular, texture regularity features such as entropy as well as the location with respect to the center of the lesion and its angular distribution turn out to be very relevant.

4.2.8 Unspecific patterns:

They are mainly described by its geometry and texture. Nevertheless, drawing conclusions about this pattern is more difficult as it includes every region that cannot be easily categorized into any of the other structures.

As mentioned before, we have complemented the set of 8 experts with a new one, referred to as *global*, which works over the whole lesion without discriminating dermoscopic structures. For this global expert, we have found that color and luminance clinical symmetry features are great indicators of malignant lesions, which usually present an asymmetrical distribution. In addition, some of the texture features such as the ring- and circular sector-based features and those measuring texture regularity, are also representative of this asymmetry and diversity of patterns, as shown in Figure 8. Furthermore, the

color-naming descriptor by van der Weijer et *al* [39] has also resulted effective to model the color distribution in this case since a more uniform distribution involves the presence of a greater variety of colors, which could be also a sign of malignancy.

4.3 Automatic melanoma diagnosis

We have evaluated the performance of our *structure-based* approach and compared it with three methods in the literature. In particular, we have implemented a CAD system relying on the clinical-based features proposed by Rubegni et al. in [32], but using a SVM for classification (instead of the simpler linear Bayes model of the original method). Moreover, we have considered a second system that complements the original feature set in [32] with new features based on general texture filter-banks (LoG, Gabor). Finally, we have also implemented a third reference system based on the dermoscopy *7-point checklist* (7PCL) [5], which is a scoring-based automatic diagnosis method. For that purpose, we have trained individual classifiers for the structures involved in the seven criteria to make binary decisions (typical vs. atypical) on these structures. Then, for each lesion in the test set, we provide a score calculated by thresholding the outputs of the classifiers, and use the corresponding score as described in the original method [5].

The results obtained for the compared methods in the two databases are summarized in Table 6. As can be seen, our *structure-based* system notably outperforms the rest of methods in both databases. The 7PCL method offers the lowest performance likely because of the thresholding and posterior discretization of the classifier outputs which, although are easy to carry out by a non-expert physician following simple rules, are far from being optimal when compared with the other methods. It can be concluded that using a system that detects and characterizes the dermoscopic structures typically used by specialists in their diagnosis works considerably better than other approaches based on global features extracted from the whole lesion or the widely used 7PCL criteria, which bases its diagnosis on a simple sum of scores.

Furthermore, we have included two versions of our approach: one that only uses the global classifier and neglects the dermoscopic structures and the final one, which includes all the information (8 dermoscopic experts and the global classifier). As shown in the table, a notable increase in the performance is achieved when all the experts are considered for the diagnosis. This is a nice result as it demonstrates the usefulness of the dermoscopic features and validates our approach for weak segmentation using image-level annotations.

Moreover, in the case of the CDI-IMO database, we have also included as a reference the score obtained by the diagnosis of a dermatologist expert in dermoscopy, who annotated the dataset. As can be observed, the performance of our approach is quite close to that provided by the expert, which strongly supports our development. Indeed, the performance obtained is even more remarkable when considering that the dermatologist may also rely on additional

information available for diagnosis, such as that in the clinical history of the patients (age, antecedents, etc.). Finally, the last row in in Table 6 shows the result of linear combination of both the automatic and manual diagnosis. As we can see, the combination of our system with the diagnosis of the dermatologist strongly boosts the performance and surpasses the ability of an expert in dermoscopy. This last result demonstrates the usefulness of our approach for Computer Aided Diagnosis and its potential to aid dermatologists in their daily practice.

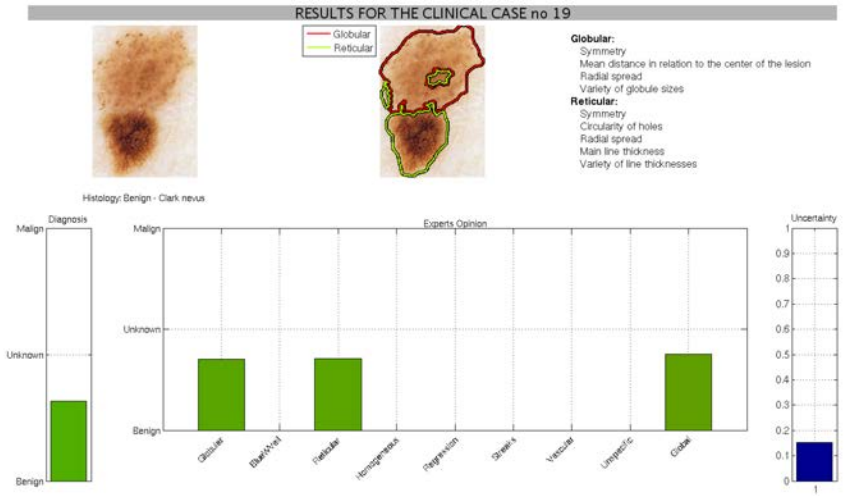
Table 6: Classification Results in AUC

Method	**EDRA [6]**	**CDI-IMO**
Features in [32]	0,788	0,658
Features in [32] + Texture Filterbanks	0,816	0,665
7PCL	0,759	0,593
Our approach (global classifier)	0,850	0,757
Our approach (all dermoscopic structures)	**0,875**	**0,792**
Dermoscopy Expert		**0,799**
Our approach + Dermoscopy Expert		**0,846**

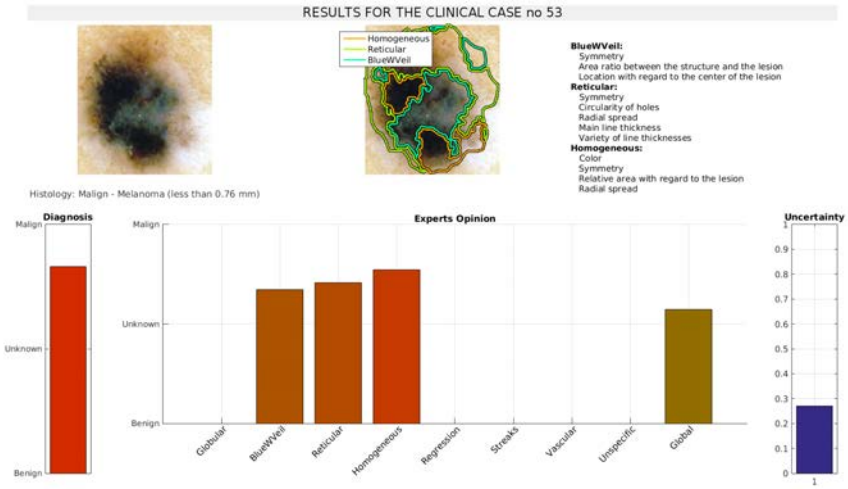
4.4 Comparison with the state of the art

An additional dataset from the ISBI 2016 challenge (Skin Lesion Analysis Towards Melanoma Detection) [16] has been used to assess the performance of our approach in comparison with current state of the art methods. The organizers released a dataset of annotated pigmented skin lesions from the ISIC archive. It gathers dermoscopic images coming from the daily clinical practice of a wide range of medical centers. This variety of sources implies that the acquisition devices are different too, so as the image sizes, which vary between 640x480 and 1944x2592. The dataset was split into training set (900 images) and test set (379 images), a division that has been preserved for a fair comparison with the systems in the leaderboard. To do so, the 25th, median and 75th percentiles of the AP values of the official submissions has been provided in Table 7 together with the AP values of the top five submissions. The evaluation measures have been obtained from the website of the challenge[3]. Performances are ordered using the AP value, as it was the chosen measure by the challenge organizers. The result of our approach is computed using the same test set that was given in the challenge, and it is trained with the released training images. Both sets are enlarged through data augmentation, in order to test the performance of the system with a large set of images [21].

[3] https://challenge.kitware.com/#phase/5667455bcad3a56fac786791



(a)



(b)

Fig. 11: Output of the system for two clinical cases: a) a benign lesion, and b) a malignant lesion

As shown in Table 7, our approach compares well with the official submissions, achieving an AP that would have been ranked in 4th position out of 25 competitors. As discussed in the final paper written by the challenge organizers [16], the predominant winning methods used Convolutional Neural Networks in their submissions, a paradigm that has become prevalent for many tasks in computer vision. Although our method does not reach the performance of the best 3 CNNs, it additionally provides very useful information

Table 7: Results obtained from our approach compared to the top 5 submissions, and some other statistical percentiles of the official Challenge ISBI 2016 leaderboard in lesion classification.

Rank	Author	AP
#1	Peter Zheng	0.637
#2	Sahbi Chaiev	0.619
#3	Balde Thierno	0.598
(#4)	**Our approach**	0.567
#4	Andre Esteva	0.563
#5	Jordan Yap	0.559
-	25th percentile	0.552
-	Median	0.494
-	75th percentile	0.347

for dermatologists that may help them in the daily practice. This last point is probably the main drawback of deep learning, which usually becomes a black box for users, being hard to interpret, and preventing from a broad adoption of CAD systems by the clinicians.

4.5 Potential applications: supporting decisions and training new specialists

As mentioned in the introduction, one of the goals of our approach, is not only to provide an alternative diagnosis that might be of help for specialists in their daily practice, but also to complement that diagnosis with extra and valuable information supporting it. For that reason, during the design of the system, we have focused on the interpretability of the data at every module of the system, for example: the segmentation into dermoscopic structures allows to the provide the physician with the dermoscopic structures detected in the lesion and, furthermore, this segmentation becomes the basis for the subsequent diagnosis using several independent classifiers; individual experts are based on features which are inspired in medical insights; the Bayesian method, which fuses the outputs of the individual experts, calculates various metrics that characterize the clinical case (global uncertainty and individual experts' opinion); etc.

In order to take advantage of all this information, we have designed an interface that presents graphically the information provided by our CAD system for each clinical case in an appealing way. The interface is illustrated in Figure 11 for two cases. As can be seen, the system not only provides a soft probability regarding its final decision, but also provides the uncertainty associated with that decision, and additional valuable information such as the segmentation of the lesion into dermoscopic structures of interest, the influence of each structural pattern expert on the final decision, and the most relevant features associated with each structure.

From our point of view, this tool can be of great interest for both expert and non-expert clinicians. For dermatologists with more expertise on dermoscopy, the tool provides extra information to support them on their decisions and

daily practice. For non-expert clinicians, the system provides a complete set of previous cases carefully analyzed which can be used as a very useful training tool, giving meaningful information about a clinical case and showing those areas and properties of the lesion that are relevant for the diagnosis.

5 Conclusion

In this paper we have proposed a CAD system for early melanoma diagnosis from dermoscopic images. The main goal of the system is not only to provide a tentative and reliable diagnosis, but also to provide the dermatologists with additional information about the lesion to help them to make their decisions. Hence, the interpretability of the results of the system has become a key requisite during the design of every module.

As in other recent works, the system relies on a previous segmentation of the lesion into dermoscopic structures of interest for the dermatologists. In our case, this first segmentation has allowed us to develop a specific classifier for each structural pattern, working on a reduced set of properly selected features. The outputs of these structure-specific classifiers are finally fused by means of a Bayesian approach which offers a final diagnosis. Moreover, this Bayesian fusion stage successfully handles the problem of missing data, which appears because not every structure is always present in each lesion.

Our approach has been evaluated in two datasets, one from an atlas of dermoscopy with images showing clear examples of the dermoscopic structures, and another more challenging containing more difficult lesions, in which only few and weak evidences appear. Our results demonstrate that our method outperforms various references in the literature that follow a similar approach. In addition, it achieves very close results to those obtained by an expert dermatologist and the combination of both automatic and manual diagnosis shows a great potential notably surpassing the human performance. Furthermore, a comparison with the official submissions of the ISBI 2016 challenge in melanoma detection shows that our method is competitive, and would have been ranked in 4th position out of 25 submissions.

It is also worth mentionning that our system provides the dermatologists with additional valuable information, which allows the doctors to gain insight the reasons behind the system decision and the confidence of the system in it. This high degree of interpretability of the system decision allows us to suggest new applications of our method, such us the training of novel doctors.

Our further lines of research include the use of novel techniques in computer vision, such as Convolutional Neural Networks (CNNs), for image segmentation and lesion classification. However, one important requisite is to maintain a high degree of interpretability, something which is not straightforward when working with deep neural networks.

Acknowledgements The authors would like to thank both *Clinica Dermatologica Internacional (CDI)* and *Instituto Madrileo de Oncologia (IMO)* for making the dermoscopy images available for research purposes.

This work has been partially supported by the National Grant TEC2014-53390-P of the Spanish Ministry of Economy and Competitiveness.

References

1. Abe, S.: Support Vector Machines for Pattern Classification (Advances in Pattern Recognition). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2005)
2. Abedini, M., Chen, Q., Codella, N., Garnavi, R., Sun, X.: chap. Accurate and Scalable System for Automatic Detection of Malignant Melanoma, pp. 293–343. Digital Imaging and Computer Vision. CRC Press (2015). DOI 10.1201/b19107-11. URL http://dx.doi.org/10.1201/b19107-11. 0
3. Abedini, M., Codella, N.C.F., Connell, J.H., Garnavi, R., Merler, M., Pankanti, S., Smith, J.R., Syeda-Mahmood, T.: A generalized framework for medical image classification and recognition. IBM Journal of Research and Development **59**(2/3), 1:1–1:18 (2015). DOI 10.1147/JRD.2015.2390017
4. Aminikhanghahi, S., Shin, S., Wang, W., Jeon, S.I., Son, S.H.: A new fuzzy gaussian mixture model (fgmm) based algorithm for mammography tumor image classification. Multimedia Tools and Applications pp. 1–15 (2016). DOI 10.1007/s11042-016-3605-x. URL http://dx.doi.org/10.1007/s11042-016-3605-x
5. Argenziano, G., Fabbrocini, G., Carli, P., De Giorgi, V., Sammarco, E., Delfino, M.: Epiluminescence Microscopy for the Diagnosis of Doubtful Melanocytic Skin Lesions. Archives of Dermatology **134**(12) (1998). DOI 10.1001/archderm.134.12.1563. URL http://dx.doi.org/10.1001/archderm.134.12.1563
6. Argenziano, G., Soyer, H.P., Giorgi, V.D.: Interactive Atlas of Dermoscopy (2002)
7. Arroyo, J., Zapirain, B.: chap. Comparison of Image Processing Techniques for Reticular Pattern Recognition in Melanoma Detection, pp. 131–181. Digital Imaging and Computer Vision. CRC Press (2015). DOI 10.1201/b19107-6. URL http://dx.doi.org/10.1201/b19107-6. 0
8. Barata, C., Celebi, M.E., Marques, J.S.: Improving dermoscopy image classification using color constancy. IEEE Journal of Biomedical and Health Informatics **19**(3), 1146–1152 (2015). DOI 10.1109/JBHI.2014.2336473
9. Barata, C., Marques, J.S., Celebi, M.E.: Improving dermoscopy image analysis using color constancy. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 3527–3531 (2014). DOI 10.1109/ICIP.2014.7025716
10. Ben Youssef, B.: A visualization tool of 3-d time-varying data for the simulation of tissue growth. Multimedia Tools and Applications **73**(3), 1795–1817 (2014). DOI 10.1007/s11042-013-1657-8. URL http://dx.doi.org/10.1007/s11042-013-1657-8
11. Benavente, R., Vanrell, M., Baldrich, R.: Parametric fuzzy sets for automatic color naming. Journal of The Optical Society of America A-optics Image Science and Vision **25** (2008). DOI 10.1364/JOSAA.25.002582
12. Berger, J.: Statistical decision theory and Bayesian analysis, 2. ed edn. Springer series in statistics. Springer, New York, NY [u.a.] (1985). URL http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+027440176&sourceid=fbw_bibsonomy
13. Codella, N., Cai, J., Abedini, M., Garnavi, R., Halpern, A., Smith, J.R.: Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images, pp. 118–126. Springer International Publishing, Cham (2015). DOI 10.1007/978-3-319-24888-2_15. URL http://dx.doi.org/10.1007/978-3-319-24888-2_15
14. Fabbrocini, G., Vita, V.D., Cacciapuoti, S., Leo, G.D., Liguori, C., Paolillo, A., Pietrosanto, A., Sommella, P.: Automatic Diagnosis of Melanoma Based on the 7-Point Checklist, pp. 71–107. Springer Berlin Heidelberg, Berlin, Heidelberg (2014). DOI 10.1007/978-3-642-39608-3_4. URL http://dx.doi.org/10.1007/978-3-642-39608-3_4
15. González-Díaz, I., Díaz-de-María, F.: A region-centered topic model for object discovery and category-based image segmentation. Pattern Recognition **46**(9), 2437–2449 (2013).

DOI 10.1016/j.patcog.2013.01.034. URL http://dx.doi.org/10.1016/j.patcog.2013.01.034

16. Gutman, D., Codella, N.C.F., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). CoRR **abs/1605.01397** (2016). URL http://arxiv.org/abs/1605.01397
17. Isasi, A.G., Zapirain, B.G., Zorrilla, A.M.: Melanomas non-invasive diagnosis application based on the abcd rule and pattern recognition image processing algorithms. Comp. in Bio. and Med. **41**(9), 742–755 (2011). URL http://dblp.uni-trier.de/db/journals/cbm/cbm41.html#IsasiZZ11
18. Kawahara, J., BenTaieb, A., Hamarneh, G.: Deep features to classify skin lesions. In: IEEE International Symposium on Biomedical Imaging (IEEE ISBI), pp. 1397–1400 (2016)
19. Khan, R., van de Weijer, J., Khan, F.S., Muselet, D., Ducottet, C., Barat, C.: Discriminative color descriptors. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pp. 2866–2873 (2013). DOI 10.1109/CVPR.2013.369
20. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (eds.) Advances in Neural Information Processing Systems 25, pp. 1097–1105. Curran Associates, Inc. (2012). URL http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf
21. L, A., R, P., P, R., et al: Digital dermoscopy analysis for the differentiation of atypical nevi and early melanoma: A new quantitative semiology. Archives of Dermatology **135**(12), 1459–1465 (1999). DOI 10.1001/archderm.135.12.1459. URL +http://dx.doi.org/10.1001/archderm.135.12.1459
22. Leo, G.D., Paolillo, A., Sommella, P., Fabbrocini, G., Rescigno, O.: A software tool for the diagnosis of melanomas. In: Instrumentation and Measurement Technology Conference (I2MTC), 2010 IEEE, pp. 886–891 (2010). DOI 10.1109/IMTC.2010.5488165
23. Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. International Journal of Computer Vision **43**(1), 29–44 (2001). DOI 10.1023/A:1011126920638. URL http://dx.doi.org/10.1023/A:1011126920638
24. Madooei, A., Drew, M.S., Sadeghi, M., Atkins, M.S.: Intrinsic Melanin and Hemoglobin Colour Components for Skin Lesion Malignancy Detection, pp. 315–322. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). DOI 10.1007/978-3-642-33415-3_39. URL http://dx.doi.org/10.1007/978-3-642-33415-3_39
25. Martnez-Corts, T., Fernndez-Torres, M.., Jimnez-Moreno, A., Gonzlez-Daz, I., de Mara, F.D., Guzmn-De-Villoria, J.A., Fernndez, P.: A bayesian model for brain tumor classification using clinical-based features. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 2779–2783 (2014). DOI 10.1109/ICIP.2014.7025562
26. Menzies, S.W., Ingvar, C., Crotty, K.A., McCarthy, W.H.: Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features. Archives of Dermatology **132**(10), 1178–1182 (1996). DOI 10.1001/archderm.1996.03890340038007. URL +http://dx.doi.org/10.1001/archderm.1996.03890340038007
27. Mirzaalian, H., Lee, T.K., Hamarneh, G.: Learning features for streak detection in dermoscopic color images using localized radial flux of principal intensity curvature. In: Mathematical Methods in Biomedical Image Analysis (MMBIA), 2012 IEEE Workshop on, pp. 97–101 (2012). DOI 10.1109/MMBIA.2012.6164758
28. Nachbar, F., Stolz, W., Merkle, T., Cognetta, A.B., Vogt, T., Landthaler, M., Bilek, P., Braun-Falco, O., Plewig, G.: The abcd rule of dermatoscopy. Journal of the American Academy of Dermatology **30**(4), 551–559 (XXXX). DOI 10.1016/S0190-9622(94)70061-3. URL http://dx.doi.org/10.1016/S0190-9622(94)70061-3
29. Pehamberger, H., Steiner, A., Wolff, K.: In vivo epiluminescence microscopy of pigmented skin lesions. i. pattern analysis of pigmented skin lesions. Journal of the American Academy of Dermatology **17**(4), 571 – 583 (1987). DOI http://dx.doi.org/10.1016/S0190-9622(87)70239-4. URL http://www.sciencedirect.com/science/article/pii/S0190962287702394

30. Pehamberger, H., Steiner, A., Wolff, K.: In vivo epiluminescence microscopy of pigmented skin lesions. i. pattern analysis of pigmented skin lesions. Journal of the American Academy of Dermatology **17**(4), 571 – 583 (1987). DOI http://dx.doi.org/10.1016/S0190-9622(87)70239-4. URL http://www.sciencedirect.com/science/article/pii/S0190962287702394
31. Rubegni, P., Cevenini, G., Burroni, M., Bono, R., Sbano, P., Biagioli, M., Risulo, M., Nami, N., Perotti, R., Miracco, C., Fimiani, M.: Objective follow-up of atypical melanocytic skin lesions: a retrospective study. Archives of Dermatological Research **302**(7), 551–560 (2010). DOI 10.1007/s00403-010-1051-6. URL http://dx.doi.org/10.1007/s00403-010-1051-6
32. Sadeghi, M., Lee, T.K., McLean, D., Lui, H., Atkins, M.S.: Global pattern analysis and classification of dermoscopic images using textons. pp. 83,144X–83,144X–6 (2012). DOI 10.1117/12.911818. URL http://dx.doi.org/10.1117/12.911818
33. Serrano, C., Acha, B.: Pattern analysis of dermoscopic images based on markov random fields. Pattern Recognition **42**(6), 1052–1057 (2009). URL http://dblp.uni-trier.de/db/journals/pr/pr42.html#SerranoA09
34. Sez, A., Serrano, C., Acha, B.: Model-based classification methods of global patterns in dermoscopic images. IEEE Transactions on Medical Imaging **33**(5), 1137–1147 (2014). DOI 10.1109/TMI.2014.2305769
35. Tanaka, T., Torii, S., Kabuta, I., Shimizu, K., Tanaka, M.: Pattern classification of nevus with texture analysis. IEEJ Transactions on Electrical and Electronic Engineering **3**(1), 143–150 (2008). DOI 10.1002/tee.20246. URL http://dx.doi.org/10.1002/tee.20246
36. Varma, M., Zisserman, A.: A statistical approach to texture classification from single images. International Journal of Computer Vision **62**(1–2), 61–81 (2005)
37. van de Weijer, J., Schmid, C.: Coloring local feature extraction. In: Proceedings of the 9th European Conference on Computer Vision - Volume Part II, ECCV'06, pp. 334–348. Springer-Verlag, Berlin, Heidelberg (2006). DOI 10.1007/11744047_26. URL http://dx.doi.org/10.1007/11744047_26
38. van de Weijer, J., Schmid, C., Verbeek, J., Larlus, D.: Learning color names for real-world applications. IEEE Transactions on Image Processing **18**(7), 1512–1523 (2009). DOI 10.1109/TIP.2009.2019809
39. Weinstock, M.A.: Cutaneous melanoma: public health approach to early detection. Dermatologic Therapy **19**(1), 26–31 (2006). DOI 10.1111/j.1529-8019.2005.00053.x. URL http://dx.doi.org/10.1111/j.1529-8019.2005.00053.x
40. Weldon, T.P., Higgins, W.E., Dunn, D.F.: Efficient gabor filter design for texture segmentation (1996)
41. Zare, H., Taghi Bahreyni Toossi, M.: chap. Early Detection of Melanoma in Dermoscopy of Skin Lesion Images by Computer Vision-Based System, pp. 345–384. Digital Imaging and Computer Vision. CRC Press (2015)
42. Zhang, M., Yang, Y., Shen, F., Zhang, H., Wang, Y.: Multi-view feature selection and classification for alzheimer's disease diagnosis. Multimedia Tools and Applications pp. 1–15 (2016). DOI 10.1007/s11042-015-3173-5. URL http://dx.doi.org/10.1007/s11042-015-3173-5
43. Zortea, M., Schopf, T.R., Thon, K., Geilhufe, M., Hindberg, K., Kirchesch, H., Mllersen, K., Schulz, J., Skrvseth, S.O., Godtliebsen, F.: Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists. Artificial Intelligence in Medicine **60**(1), 13 – 26 (2014). DOI http://dx.doi.org/10.1016/j.artmed.2013.11.006. URL http://www.sciencedirect.com/science/article/pii/S0933365713001589